

## **An automated validation and alert system for continuous environmental monitoring data**

**Kirk Barrett, Amir Mirza, Richard Holowczak**

Dr. Kirk Barrett is the Research Director of the Meadowlands Environmental Research Institute (MERI), Rutgers University CIMIC, Newark NJ (email: [kbarrett@cimic.rutgers.edu](mailto:kbarrett@cimic.rutgers.edu); web: <http://cimic.rutgers.edu/meri>). Mr. Mirza is a Research Associate with MERI. Dr. Holowczak is an Assistant Professor in the Department of Computer Information Systems, City University of New York.

The Meadowlands Environmental Research Institute's network of continuous environmental monitoring stations in the Hackensack Meadowlands, a complex, urban, estuarine system in northeastern New Jersey, measures hydrologic, water quality, weather and air quality at 5-minute intervals. A telecommunication and database management system retrieves and stores data automatically every hour and makes it available through a Worldwide Web interface in near real time. With more and more environmental monitoring data similarly collected continuously and delivered over the Internet, the need for automated, continuous validation is becoming increasingly important.

A computerized system was developed that automatically checks newly collected data every hour. The system checks the data against maximum and minimum expected values and maximum and minimum expected variability. The system flags data values that violate an expected condition with a code indicating the condition, with the codes stored in the database along with the values. Data that does not violate an expectation is also coded to indicate it has been automatically validated. A report is emailed to sensor and resource managers, alerting them to the suspicious data that may indicate that a sensor has malfunctioned and needs to be fixed, or that an extreme event was or is occurring and management action is needed. To cite a specific benefit of the system, the Meadowlands is a low-lying area subject to coastal flooding during storms. The system will alert resource managers to flooding monitored by water level sensors.

Data managers go through the data and delete flagged data, modify a code to indicate that flagged data is indeed valid (ie, manually validated data), or manually flag data with violation codes (manually invalidated data). We plan to develop and test more sophisticated validation techniques, including seasonally varying, multivariate and cross-media tests: for example, during periods of solar radiation at water temperatures above 5C, dissolved oxygen is expected to be increasing (because of photosynthesis in the water column).